

基于 MCCV 奇异样本筛选和 CARS 变量选择法对蜂蜜 pH 值和酸度的近红外光谱检测

李水芳¹, 单杨^{2,*}, 范伟³, 尹永⁴, 周孜⁴, 李高阳²

(1.中南林业科技大学理学院, 湖南 长沙 410004; 2.湖南省食品测试分析中心, 湖南 长沙 410025;

3.中南大学化学化工学院, 湖南 长沙 410083; 4.湖南明园蜂业有限公司, 湖南 长沙 410005)

摘要:采用 Norris 平滑加一阶微分数据预处理, 蒙特卡洛交互验证(MCCV)的奇异样本筛选和 CARS(competitive adaptive reweighted sampling)变量选择法, 用 Kennard-Stone(KS)法划分训练集和预测集, 偏最小二乘(PLS)回归近红外光谱建模, 对蜂蜜 pH 值和酸度进行定量分析。pH 值和酸度校正模型的交互验证决定系数(R_{cv}^2)、交互验证均方差(RMSECV)、预测集决定系数(R_p^2)、预测均方差(RMSEP)分别为 0.8516 和 0.8723、0.1214 和 2.1734、0.8205 和 0.8250、0.1196 和 2.4674。结果表明, 该方法适于蜂蜜 pH 值的测定, 而不宜用于测定蜂蜜酸度。

关键词:近红外光谱; 蒙特卡洛交互验证的奇异样本筛选; CARS 变量选择; 蜂蜜; pH 值; 酸度

Analysis of pH and Acidity of Honey by Near Infrared Spectroscopy Based on MCCV Outlier Detection and CARS Variable Selection

LI Shui-fang¹, SHAN Yang^{2,*}, FAN Wei³, YIN Yong⁴, ZHOU Zi⁴, LI Gao-yang²

(1. College of Science, Central South University of Forestry and Technology, Changsha 410004, China; 2. Hunan Food Test and Analysis Center, Changsha 410025, China; 3. School of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China; 4. Hunan Mingyuan Honey Products Co.Ltd., Changsha 410005, China)

Abstract: The near infrared spectra of honey samples were calculated by the method of Norris smoothing combined with first derivative. The outliers were detected by Monte Carlo cross validation (MCCV), and the variables were selected by competitive adaptive reweighted sampling (CARS). The samples were divided into calibration set and validation set by Kennard-Stone (KS) algorithm. Partial least squares (PLS) regression was applied to build a quantitative calibration model of pH and acidity. The coefficient of cross-validation of the calibration set (R_{cv}^2) was 0.8516, and the root mean square error of cross-validation (RMSECV) was 0.1214. The coefficient of determination of the validation set (R_p^2) was 0.8205, and the root mean square error of prediction (RMSEP) was 0.1196 for pH value. For acidity, the R_{cv}^2 , RMSECV, R_p^2 and RMSEP were 0.8723, 2.1734, 0.8250 and 2.4674, respectively. The finding shows that this method is suitable for quantitative analysis of honey pH, while caution is needed for honey acidity analysis.

Key words: near infrared spectroscopy; outliers detection of Monte Carlo cross validation (MCCV); competitive adaptive reweighted sampling for variable selection; honey; pH value; acidity

中图分类号: TS207.3

文献标识码: A

文章编号: 1002-6630(2011)08-0182-04

蜂蜜是蜜蜂采集植物的花蜜、分泌物或蜜露, 与自身分泌物结合后, 经充分酿造而成的天然甜味物质, 主要成分为葡萄糖和果糖, 此外, 还含有糊精、蛋白质、蜡质、氨基酸、矿物质、芳香化合物、有机酸、色素、挥发油、酶类、维生素、花粉、类黄酮、羟

甲基糠醛等多种成分^[1-2], 是一种深受广大消费者喜爱的天然保健食品。

pH 值是描述氢离子活度的参数。蜂蜜的 pH 值可作为区分花蜜和甘露蜜^[3]以及判断单花蜜是否真实^[4]的一个重要参数。酸度一般是指中和 100g 试样中的酸性物质所

收稿日期: 2010-05-29

基金项目: “十一五” 国家科技支撑计划项目(2009BADB9B07)

作者简介: 李水芳(1968—), 女, 副教授, 硕士, 研究方向为森林食物资源的开发与利用。

E-mail: csfulishuifang@126.com

* 通信作者: 单杨(1963—), 男, 研究员, 博士, 研究方向为农产品安全检测。E-mail: sy6302@163.com

需要的氢氧根离子(OH^-)的毫摩尔数。蜂蜜中的酸性物质包括有机酸和无机酸,有机酸有葡萄糖酸、乳酸、醋酸、丁酸、甲酸和苯甲酸等,无机酸有磷酸和盐酸等^[5]。蜂蜜酸度反映了蜂蜜中酸性物质含量的高低,是衡量蜂蜜是否发酵的重要指标^[6],也可作为区分花蜜和甘露蜜的一个重要参数^[3]。我国蜂蜜国家标准规定蜂蜜酸度不得大于 40mL/kg (1mol/L 氢氧化钠)^[6]。

近红外光谱(near infrared spectroscopy, NIR)技术是一种快速无损的检测技术,能实现对物质多种理化指标的快速无损分析,已有文献报道该技术可用于测定食品的 pH 值和酸度^[7-9],蜂蜜中糖类成分的 NIR 检测已有报道^[10-11],而蜂蜜 pH 值和酸度的 NIR 检测国内还未见报道,国外虽有报道^[12-13],但模型预测结果差。本实验利用 NIR 技术结合基于蒙特卡洛交互验证(Monte-Carlo cross-validation, MCCV)的奇异样本筛选法^[14]和 CARS (competitive adaptive reweighted sampling)变量选择法^[15],对蜂蜜 pH 值和酸度进行分析,以期实现 NIR 检测,为快速评价蜂蜜质量提供一定的参考依据。

1 材料与方法

1.1 样品及处理

样本集包括真假蜂蜜样本共 134 个。真蜂蜜样本包括 2008 年和 2009 年收集的 15 个品种(油菜、柑橘、荔枝、洋槐、荆条、枸杞、益母草、野菊花、九龙藤、黄芪、桂圆、枣花、山桂花、老瓜头、苹果),共 122 个,来自全国 10 个省区(湖南、湖北、广西、四川、宁夏、贵州、山西、陕西、河南、云南)。掺假蜂蜜样本分别用果葡糖浆和麦芽糖浆按不同比例掺假得到,每种掺假物得到 6 个掺假样本,共 12 个。蜂蜜收集后,贮藏在 $6\sim 8^\circ\text{C}$ 冰箱中。采集光谱时,先将蜂蜜从冰箱内取出,放置至室温,再在 55°C 水浴下溶解晶体。

1.2 仪器与设备

Antaris II 傅里叶变换近红外光谱仪(带 TQ 软件) 美国 Thermo Nicolet 公司; PHS-320 型 pH 计 成都世纪方舟科技有限公司。

1.3 方法

1.3.1 近红外光谱条件

扫描范围 $10000\sim 4000\text{cm}^{-1}$,分辨率为 8cm^{-1} ,增益 2,扫描次数 32 次。

1.3.2 pH 值

采用 pH 计测定。平行测定 3 次,取平均值。

1.3.3 酸度

对颜色较深的蜂蜜,参照国标方法^[16]测定并做一定修改:准确称取 10g 左右蜂蜜溶于 100mL 蒸馏水中,用

0.05mol/L NaOH 溶液滴定到 pH8.3,记录消耗的 NaOH 溶液体积(mL),再换算为中和 1000g 蜂蜜所需消耗的 1mol/L NaOH 溶液毫升数。平行 3 次测定,取平均值。

1.3.4 光谱采集

先将样品搅拌均匀,再置于离心试管中离心 10min 除去气泡,取约 10g 样品倒入一专用玻璃容器内,用金箔控制样品厚度为 0.1mm,以空气为背景,每小时扫描一次背景,采用透反射方式在约 25°C 下采集光谱。取 4 次的平均光谱为该样品的原始光谱。每采集完 1 个样品,先用自来水洗净玻璃容器和金箔,再用蒸馏水润洗两遍,然后冷风吹干。光谱采集完成后即进行 pH 值和酸度测定。同一批样品,光谱采集和数据测定在同一天进行。

2 结果与分析

2.1 原始光谱和预处理光谱

平滑可提高分析信号的信噪比,微分能有效消除光谱漂移,二者都是常用的光谱预处理方法。本实验通过比较 8 种光谱预处理方法的效果,分别为原始光谱、1 阶微分(first derivative, 1D)、2 阶微分(second derivative, 2D)、Savitzky-Golay(S-G)平滑、Norris 平滑+1D、Norris 平滑+2D、S-G 平滑+1D、S-G 平滑+2D,最终选择 Norris 平滑+1D 为本实验的最优光谱预处理方法。图 1(A)为样本原始光谱,存在较明显的基线漂移。图 1(B)为(Norris 平滑+1D)预处理后的光谱,光谱经平滑和微分处理后,能有效降低基线漂移,消除光谱变量间的差异。

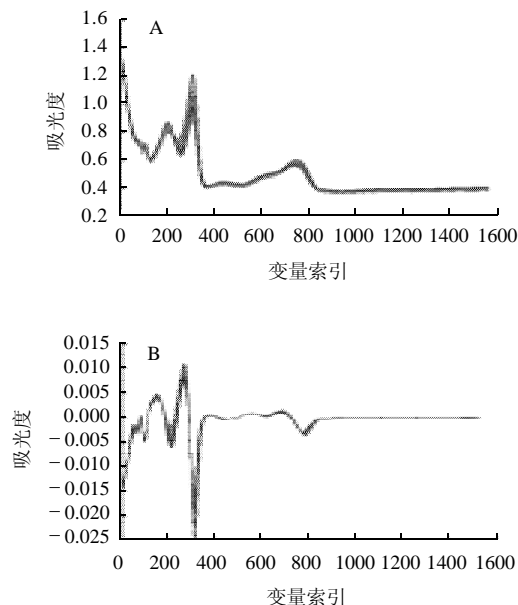


图1 样本原始光谱图(A)和(Norris 平滑+1D)预处理后的光谱图(B)
Fig.1 Original (A) and (Norris smoothing + 1D) pretreated (B) spectra of samples

2.2 MCCV 的奇异样本的剔除

MCCV 的奇异样本筛选法是一种新近提出的筛选奇异样本的方法,能够在一定程度上降低由掩蔽效应带来的风险,有效检出光谱阵和性质阵方向的奇异点,与传统方法相比具有较高的识别奇异样本的能力。用 PLS 确定最佳主成分数后,用蒙特卡洛随机取样(Monte Carlo sampling, MCS)法取 80% 的样本作校正集建立 PLS 回归模型,剩余部分作预测集,循环 2000 次,得到各样本的一组预测残差,求各样本预测残差的均值(MEAN)与方差(STD),作 MEAN-STD 图(图 2)。

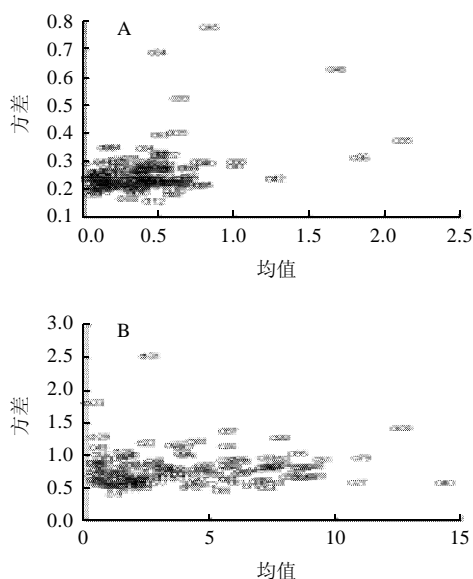


图 2 pH 值(A)和酸度(B)的均值-方差分布图

Fig.2 Mean value and variance distribution of pH value (A) and acidity (B)

从图 2 可知,某些样本明显偏离主体样本,如 A 中的 33、34、35、36、45 和 46, B 中的 27、40、79、109、123 和 131,这些样本可视为奇异样本,应剔除。从表 1 可知,奇异样本剔除前后 PLS 校正模型的交互验证均方差(RMSECV)的变化情况:pH 值的 RMSECV 由 0.3545 降到 0.1854,降低显著,酸度的 RMSECV 由 3.1779 降到 3.0891。

2.3 CARS 法选择变量

CARS 变量选择法是最新提出的变量筛选方法,可有效选择与所测性质相关的最优变量组合,预测结果往往优于 MC-UVE、全谱偏最小二乘法(FS-PLS)和移动窗口偏最小二乘法(MW-PLS)^[11]。用 CARS 变量选择法(CARS 变量选择的 Matlab 源代码获取自 <http://code.google.com/p/carspls/downloads/list>)筛选出的 pH 值和酸度的最优变量子集的变量个数都为 59 个。从表 1 可知,变量选择前后 pH 值和酸度 PLS 校正模型的交互检验结

果:交互验证决定系数(R_{cv}^2)分别由 0.6404 增大到 0.8563 (pH)和 0.7399 增大到 0.8714(酸度);RMSECV 分别由 0.1854 降低到 0.1172(pH)和 3.0891 降低到 2.1718(酸度)。

表 1 奇异样本剔除前后和变量选择前后的 PLS 校正模型的交互检验结果

Table 1 Results of cross-calibration of PLS models

处理方法	pH		酸度	
	R_{cv}^2	RMSECV	R_{cv}^2	RMSECV
Norris + 1D	0.7237	0.3545	0.7445	3.1779
Norris + 1D + MCCV	0.6404	0.1854	0.7399	3.0891
Norris + 1D + MCCV + CARS	0.8563	0.1172	0.8714	2.1718

注:1D.一阶微分;MCCV.基于蒙特卡洛交互验证的奇异样本筛选;CARS. CARS 法筛选变量; R_{cv}^2 .交互验证决定系数;RMSECV.交互验证误差均方根。下同。

2.4 模型的建立与评价

pH 值的范围为 3.18~4.80,平均值为 3.65,酸度的范围为 2.70~32.6mL/kg,平均值为 16.8mL/kg。采用 Norris 平滑 + 1D 的光谱预处理方法,剔除奇异样本,用选择的最优变量子集,通过 Kennard-Stone(KS)法划分训练集和预测集,所建立的 PLS 校正模型的预测结果见表 2,真实值与预测值之间的相关图见图 3。

表 2 pH 值和酸度校正模型的预测结果

Table 2 Forecasting results of models of pH value and acidity

	样本 总数	校正集			预测集		
		样本数	R_{cv}^2	RMSECV	样本数	R_p^2	RMSEP
pH 值	128	100	0.8516	0.1214	28	0.8205	0.1196
酸度	125	100	0.8723	2.1734	25	0.8250	2.4674

注: R_p^2 预测集决定系数;RMSEP 预测误差均方根。

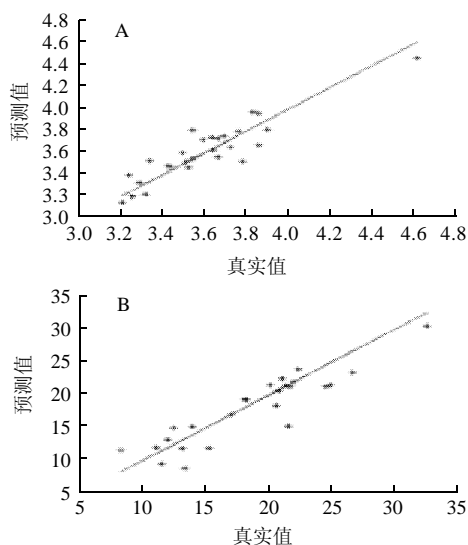


图 3 pH 值(A)和酸度(B)真实值与近红外光谱预测值相关图

Fig.3 Relationships between real and predicted values for pH (A) and acidity (B)

表3列出了相关文献报道的蜂蜜pH值和酸度的NIR模型检验结果,表中数据表明本实验模型预测结果最好。

表3 相关文献报道的蜂蜜pH值和酸度的NIR定量校正模型检验结果

Table 3 Previously reported NIR models of pH and acidity of honey and those obtained from this study

	pH值				酸度			
	校正集		预测集		校正集		预测集	
	R_{cv}^2	RMSECV	R_p^2	RMSEP	R_{cv}^2	RMSECV	R_p^2	RMSEP
文献[12]	0.622	0.3	0.657	0.3	0.636	5	0.737	4
文献[13]	0.7	0.21	—	—	—	—	—	—
本实验	0.8516	0.1214	0.8205	0.1196	0.8723	2.1734	0.8250	2.4674

注:—,文献未标出。

3 结 论

本实验利用NIR技术结合新近开发的MCCV的奇异样本筛选法和CARS变量选择法,对蜂蜜pH值和酸度进行了分析,比较了奇异样本筛选前后和变量选择前后的PLS模型预测结果,剔除奇异样本后,模型的预测结果有了显著提高,用筛选出的变量建立的模型的预测结果也优于用全部变量建模的预测结果。pH值PLS近红外光谱最优校正模型的 R_{cv}^2 为0.8516、RMSECV为0.1214,用该模型对验证集样本进行预测,预测集 R_p^2 为0.8205、RMSEP为0.1196,表明该模型可用于蜂蜜pH值的测定。酸度PLS近红外光谱最优校正模型的 R_{cv}^2 为0.8723、RMSECV为2.1734,当用该模型对验证集样本进行预测时,预测集 R_p^2 为0.8250、RMSEP为2.4674,表明本实验所建立酸度模型还不宜用于蜂蜜酸度的测定。

参考文献:

- [1] 侯瑞丽. 蜂蜜主要成分的近红外光谱检测技术的研究[D]. 沈阳: 沈阳农业大学, 2007.
- [2] DOWNEY G, FOURATIER V, KELLY J D. Detection of honey adulteration by addition of fructose and glucose using near-infrared transreflectance spectroscopy[J]. Journal of Near Infrared Spectroscopy, 2003, 11(6): 447-456.
- [3] SANZ M L, GONZALEZ M, de LORENZO C, et al. A contribution to the differentiation between nectar honey and honeydew honey[J]. Food Chemistry, 2005, 91(2): 313-317.
- [4] BENEDETTI S, MANNINO S, SABATINI A G, et al. Electronic nose and neural network use for the classification of honey[J]. Apidologie, 2004, 35(4): 397-402.
- [5] 董捷. 蜂产品加工技术[M]. 北京: 中国农业出版社, 2003: 29-31.
- [6] GB 18796—2005 蜂蜜[S]. 2005.
- [7] SHAO Yongni, HE Yong. Measurement of soluble solids and pH of Yogurt using visible/near infrared spectroscopy and chemometrics[J]. Food Bioprocess Technol, 2009, 2(2): 229-233.
- [8] TSUCHIKAWA S, HAMADA T J. Application of time-of-flight detecting sugar and acid contents in apples[J]. Agric Food Chem, 2004, 52(9): 2434-2439.
- [9] 廖宜涛, 樊玉霞, 伍学千, 等. 猪肉pH值的可见近红外光谱研究[J]. 光谱学与光谱分析, 2010, 30(3): 681-684.
- [10] 陈兰珍, 薛晓峰, 陈芳, 等. 蜂蜜中还原糖组分测定的近红外光谱应用研究[J]. 食品科学, 2009, 30(8): 147-150.
- [11] 屠振华, 朱大洲, 籍保平, 等. 蜂蜜中果糖和葡萄糖近红外检测的差异性分析与特征研究[J]. 分析化学, 2010, 38(1): 45-50.
- [12] RUOFF K, LUGINBUHL W, BOGDANOV S, et al. Quantitative determination of physical and chemical measurands in honey by near-infrared spectrometry[J]. Eur Food Res Technol, 2007, 225(3/4): 415-423.
- [13] COZZOLINO D, CORBELL A E. Determination of honey quality components by near infrared reflectance spectroscopy[J]. Journal of Apicultural Research, 2003, 42(1/2): 16-20.
- [14] 刘智超, 蔡文生, 邵学广. 蒙特卡罗交叉验证用于近红外光谱奇异样本的识别[J]. 中国科学, B辑: 化学, 2008, 38(4): 316-323.
- [15] LI Hongdong, LIANG Yizeng, XU Qingsong, et al. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration[J]. Analytica Chimica Acta, 2009, 648(1): 77-84.
- [16] SN/T 0852—2000 进出口蜂蜜检验方法[S].